Australia's National
Science Agency

Inquiry into the adoption and use of artificial intelligence (AI) systems and processes by public sector entities

Joint Standing Committee of Public Accounts and Audit

CSIRO Submission 24/102

[October 2024]

Enquiries should be addressed to:
Karen O'Rourke
CSIRO Ministerial Liaison Office
GPO Box 1700 Canberra 2601
T [redacted]
E [redacted]

Contents

Introduction

CSIRO welcomes the opportunity to provide input to the Joint Standing Committee of Public Accounts and Audit inquiry into the adoption and use of artificial intelligence (AI) systems and processes by public sector entities. As AI technology rapidly advances, CSIRO fully supports the Committee's focus on ensuring that AI adoption and use in the public sector is carried out safely and responsibly. We hope this submission provides valuable insights into how AI is being used, managed, and governed within CSIRO.

CSIRO engages with AI in various ways:

- We conduct research into the responsible and ethical use of AI,
- We use AI broadly across science domains to accelerate scientific discovery, and
- Generative AI (Gen AI) is increasingly being used as an enterprise tool to support employee productivity.

CSIRO has significant and deep experience in the use of AI in science. We are learning how to get the best out of enterprise AI tools, such as Office 365, and at the same time uplifting governance processes, monitoring and weighing the benefits and risks associated with the technology.

This response addresses all ToRs with the exception of ToR 6. ToR 6 pertains to questions about the public sector's capability to effectively adopt AI, which CSIRO does not have the relevant expertise to comment on.

CSIRO response to the Terms of Reference (ToR)

1. The purposes for which AI is currently being used by the public sector entity and whether there are planned or likely future uses

CSIRO uniquely combines cutting-edge capability in technologies such as AI with deep domain expertise in areas from healthcare to the environment and agriculture. CSIRO is using AI to tackle challenges from bushfire management to boosting agricultural productivity and protecting our environments such as the Great Barrier Reef. CSIRO's AI research includes contributions to inclusive, ethical, safe and secure AI adoption in Australia as well as robotics and autonomous systems research.

This multidisciplinary approach has allowed CSIRO to create AI solutions, both independently and in collaboration with public and private sector entities, including:

- **Spark:** A wildfire simulation toolkit for researchers and experts in the disaster resilience field which uses AI in combination with other technologies to predict the path of bushfires.
- **Responsible and safe AI (Government):** CSIRO works closely with the Australian government to provide scientific and technical advice which informs the development of responsible and safe AI policies. This includes providing technical advice on AI safety through the National AI Centre and the AI Safety Research Network and developing responsible AI best practice catalogues and AI diversity and inclusion guidelines for Australian industry. The latter has contributed to the development of the National Framework for AI Assurance for government use of AI. CSIRO has also contributed to the development of Australia's AI safety standards, supported government participation in international AI safety summits, and facilitated international research alliances, positioning Australia as a leader in responsible AI.
- **Responsible and safe AI (Industry):** CSIRO's engagement with industry on responsible and safe AI spans from startups to large organisations. Startups such as TrueRecognition and Reejig have utilised CSIRO's responsible AI best practices to enhance privacy, accuracy, and trustworthiness in their AI products, while larger firms such as Westpac and SEEK have tailored responsible AI guidelines to improve AI use in customer services and diversity efforts. CSIRO has also collaborated with the investment sector, working with the Alphinity Investment Company to integrate responsible AI into Environmental, Social, and Governance (ESG) frameworks to support broader low-emission and social goals.
- **Reef Intelligence:** Uses machine learning and artificial intelligence to identify and monitor Crown of Thorns Starfish in the Great Barrier Reef. Crown of Thorns Starfish are a major predator of corals and a major cause of coral loss. This project is in collaboration with Google with funding from the Department of Foreign Affairs and Trade.
- **Science Digital**: In collaboration with Google, CSIRO is developing AI-driven assistants for scientists across scientific disciplines, supporting their scientific discovery work from hypothesis generation to experimental planning and validation, to output analysis and insights generation.
- **Robotics:** CSIRO's use of AI and AI algorithms in robotics use cases is an example of embodied AI where AI algorithms have a physical embodiment. Examples of CSIRO's research in this space include path planning algorithms used by robots to understand their environments, navigate complex and dangerous environments (such as caves), plan trajectories in manipulation tasks using robot arms, and to interact with humans using natural language and gestures.

CSIRO provides centralised enterprise technology infrastructure to enable our science programs to use and create AI tools. In this way, CSIRO has assurance that appropriate risk management and controls are in place to mitigate potential risks in the use of AI in the organisation.

CSIRO does not currently use AI for automated decision making.

CSIRO participated in the Australian Government MS365 Copilot trial to evaluate the potential benefits for the organisation and may adopt Copilot pending final evaluation of the benefits at the closure of the trial. We have extended the trial to further validate the benefits, and better understand the risks, associated with generative AI usage and how best to manage them.

Insights from the Copilot trial have revealed mixed experiences; while some staff have reported significant productivity gains, others have struggled to adopt the technology. CSIRO is optimistic that the benefits of generative AI will grow as the technology evolves and as organisations refine their approaches to integrating it effectively.

CSIRO will continue to seek opportunities to leverage emerging AI technology in a safe and responsible way.

2. The existing legislative, regulatory and policy frameworks that are relevant to the use of AI and whether they are fit for purpose.

CSIRO follows the legislative and policy frameworks listed below. The current frameworks provide a good structure for managing AI responsibly at present. However, we may find that further work is required on the detailed technical methods and standards tailored to evolving forms of AI, especially Gen AI and frontier AI. The science on the technical assurance side is far from mature and settled, even for the current AI.

- Responsible AI in Government Policy (2024, Digital Transformation Agency - DTA)
- Voluntary AI Safety Standard (2024, Department of Industry, Science and Resources)
- AI Assurance Framework (2024, Department of Finance)
- Australian Government Interim guidance on government use of public generative AI tools (November 2023, DTA)
- Australian Code for the Responsible Conduct of Research (2018, National Health and Medical Research Council)
- *Privacy Act 1988* (Cth)

Risks to CSIRO Technology environments, including AI, are considered by CSIRO's internal responsible Director in line with the guidance of the Protective Security Policy Framework and Hosting Certification Framework, provided by the Department of Home Affairs; and the Information Security Manual, published by Australian Signals Directorate. Whilst this guidance is not specific to AI technology, CSIRO considers that these are still highly relevant and applicable to managing a number of the risks inherent with the use of AI technologies.

3. Whether the internal governance structures that currently exist for AI will ensure its ethical and responsible use by public sector entities.

CSIRO's Science and Delivery policy applies to all of CSIRO's research activities, which are conducted in line with the *Australian Code for the Responsible Conduct of Research 2018*. CSIRO has existing internal procedures and guidelines to support the responsible use of AI. Guidance on the usage of Gen AI tools has been provided to staff. CSIRO has also established an internal Responsible AI Working Group to develop an AI procedure and risk assessment process. This group monitors alignment with CSIRO's internal policies (e.g. privacy, code of conduct) and evolving government advice, including the DTA's Responsible AI in Government Policy released in September 2024.

Existing CSIRO procedures that apply to AI use include:

- Cyber Security Principles
- Responsible use of ICT and Internet services
- Software Acquisition and Management
- Record Keeping Principles

- Privacy
- Ethical Human Research – the *National Statement on Ethical Conduct in Human Research 2018*
- Indigenous Cultural and Intellectual Property Principles.

CSIRO's internal governance structures, policies, and procedures are designed to guide employees on the ethical and responsible use of AI. It is important to acknowledge that no framework can completely eliminate risk. CSIRO has safeguards in place to mitigate potential issues, recognising that AI systems are complex and constantly evolving. Therefore, we continuously monitor and refine our governance structures to address emerging risks and maintain ongoing accountability. This requires a balance between implementing robust risk mitigation strategies and being realistic about the limits of control in such a dynamic space.

It is also important to note that overly-rigid governance structures could inhibit our ability to test and adopt emerging technologies effectively. We are working to understand where the guardrails are best placed as new government advice and advancements in AI become available, while prioritising safe, responsible, and ethical AI use.

4. The internal framework/policies or additional controls used for assessing the risks associated with the use and possible misuse of AI, including the areas of security, privacy, ethics, bias, discrimination, transparency and accountability.

CSIRO applies a standard project-based approach to enterprise technology adoption. Risks and issues are identified and managed via our standard project management framework. Risks and controls are routinely reviewed and updated in response to emerging risks, including AI use cases. Additionally, technology risks and controls are considered and adopted in line with guidance and frameworks including the Protective Security Policy Framework, the Information Security Manual, and Hosting Certification Framework. Given the emerging nature and rapid cycle of change in AI technologies entering the market, CSIRO also regularly reviews new guidance when it is released and adopts relevant and appropriate recommendations, for example the Australian Cyber Security Centre (ACSC's) recent guidance on "Engaging with artificial intelligence".

Any new technologies sourced by CSIRO are procured in accordance with our obligations under the *Public Governance, Performance and Accountability Act 2013 (Cth)* (PGPA Act). Vendor assurance is supplied via our internal cyber security teams, and our contract and governance teams. Supply chain risks are considered as part of this evaluation process. Emerging risks, including AI supply chain risks, are monitored by the internal Responsible AI Working Group and communicated to inform decision making.

All third-party software updates are assessed prior to introduction to the CSIRO enterprise environment, in line with enterprise risk management policies and standard release and change management processes.

The Responsible AI Working Group is adapting the NSW AI Assessment Framework to assess risk and align the use of AI to CSIRO's AI principles.

5. Whether there is an adequate line of sight to the output of AI, and the decisions made through its use.

CSIRO's guidance on the responsible use of AI recommends employees cite the use of AI. Guidance also includes verifying any outputs generated to check for reliability, quality, or errors.

CSIRO does not use AI for automated decision making in any enterprise systems. Any decisions incorporating AI outputs are made by a human-being, in accordance with CSIRO's Delegations Framework.

CSIRO is a signatory to and has adopted as policy the Australian Code for the Responsible Conduct of Research. The code makes specific recommendations about the appropriate handling of data, and application of technology in research activities. To support compliance with the code, CSIRO has established an internal network of Research Integrity Advisors who are located in the research units and their role includes additionally providing specific guidance on the responsible application of AI in research.

6. Whether the public sector has the internal capability to effectively adopt and utilise AI into the future

Formally, CSIRO is not included in the Australian Public Sector. Other than providing the information in this submission concerning CSIRO's own capacity and policies and procedures, CSIRO is not able to provide further response on this topic.

7. Whether there are sovereign capability issues to consider given that most AI tools currently used in Australia are sourced from overseas.

With respect to AI, sovereign capability includes the availability (and scalability) of high-performance computing infrastructure, secure data storage, skilled technical workers, datasets for training/adapting AI models and the ability to manage/regulate AI model use in Australia (Hajkowicz, 2024 a).

There are concerns about international market concentration and monopoly power associated with generative AI models (e.g. large language models), making it difficult for Australian firms to compete. Sovereign capability is an important consideration in the adoption of AI technology and the competitive interests of Australian workers and firms as next generation frontier/foundation AI models take hold. CSIRO reflects that there is scope to build more of these models in Australia with Australian data, which would improve data security and make the models work better in Australian context. Understanding the gaps in our sovereign capability and identifying ways to fill these gaps could contribute to Australia's own AI industry's development.

Sovereign AI capability systems also encompass data security. AI systems vacuum-up vast quantities of data. They need this data to work, but if the data is sensitive, private or confidential, this can cause concerns for citizens or governments about whether the data is secure. If the models are built (trained/tuned) and operated within Australia, then data sovereignty is improved. It is important to address the need for data sovereignty and ensure Australian data is stored and processed within national borders when necessary (data locality).

There's also the matter of whether AI systems trained on data in other countries will work well for Australia in fields like healthcare, weather/climate forecasting, traffic management and many other critical areas. Australian data is likely to make AI systems perform better (sometimes just work) in Australian applications. Basically, if we train AI on Australian data there will be better outcomes for Australian industry and society.

Indigenous data sovereignty is also a key consideration. Global AI models are absorbing data created by indigenous Australians (e.g. artworks) making it possible for these artworks to be replicated without knowledge, consent or payment to the original indigenous creators. Furthermore, there is no guarantee that the outputs of AI models built by global tech corporations will accurately convey Australian indigenous culture and they may not feedback the benefits to indigenous communities from the data they use (Carlson, 2023). Indigenous communities are already, and can increasingly, drive Australian-made AI model development.

Sovereign capability matters are being actively examined in other countries facing similar challenges. For example, the UK Competition and Markets Authority (CMA, 2024) has studied the rising use of overseas sourced AI foundation models from global technology corporations such as ChatGPT within the UK economy. The CMA identifies the risk that "a small number of incumbent technology firms, which already hold positions of market power in many of today's most important digital markets, could profoundly shape FM [AI foundation model]-related markets to the detriment of fair, open and effective competition, ultimately harming businesses and consumers, for example by reducing choice and quality, and by raising prices"). They are exploring policy and regulatory interventions to ensure a fair and competitive marketplace. The Brookings Institution, a nonprofit public policy organisation based in Washington, DC, United States, also notes the risks of market concentration indicating natural monopolies are likely to form and that actions by regulators may be needed to protect smaller companies and consumers (Vipra, 2023).

CSIRO recently released a report on AI foundation models (Hajkowicz, 2024 b) which identifies sovereign capability risks/issues associated with the rising use (and dependence) on a small number of powerful AI foundation models made and operated by overseas technology corporations. This report identifies issues

relating to content moderation, data security, pricing and the competitive interests of Australian firms and workers impacted by AI foundation models. The report also notes the considerable productivity benefits and broader social benefits of AI foundation models for Australia. Identifying the right balance of policies and regulations to achieve beneficial outcomes whilst mitigating risks and ensuring sovereign capability requirements are met will be challenging.

References

Hajkowicz, S (2024 a) Artificial intelligence foundation models: Industry enablement, productivity growth, policy levers and sovereign capability considerations for Australia. CSIRO. Canberra.

Hajkowicz SA (2024 b) Artificial intelligence foundation models: Industry enablement, productivity growth, policy levers and sovereign capability considerations for Australia. CSIRO, Brisbane.

Carlson B, Richards P (2023) Indigenous knowledges informing 'machine learning' could prevent stolen art and other culturally unsafe AI practices. The Conversation (8 September).

CMA (2024) AI Foundation Models Update paper (11 April). Competition and Markets Authority, United Kingdom Government, London.Vipra J, Korinek A (2023) Market concentration implications of foundation models: The invisible hand of ChatGPT. Brookings Institution, Washington DC.

As Australia's national science agency and innovation catalyst, CSIRO is solving the greatest challenges through innovative science and technology.

CSIRO. Unlocking a better future for everyone.

www.csiro.au